Probeless Illumination Estimation for Outdoor Augmented Reality

Madsen, Claus B.; Lal, Brajesh Behari

Published in:
Augmented Reality

Publication date:
2010

Document Version
Publisher's PDF, also known as Version of record

2

Probeless Illumination Estimation for Outdoor Augmented Reality

Madsen and Lal
Aalborg University
Denmark

1. Introduction

Without doubt Augmented Reality (AR) has the potential to become a truely widespread technology. The mere concept of AR is immensely fascinating for a tremendous range of application areas, from surgery, over equipment maintenance, to more entertaining and educational applications. What drives most of the research in AR is the vision that one day we will be able to create real-time, interactive visual illusions as convincingly as we have become accostumed to seing in large Hollywood productions.
Unfortunately, there are some technical challenges in reaching this goal. Generally, there are three major challenges associated with AR: 1) camera tracking (matching position and orientation of the camera to the coordinate system of the scene), 2) handling occlusions (having sufficient 3D information of the real scene to handle occlusion between real and virtual geometry), and 3) illumination consistency (having sufficient knowledge of the real scene illumination to be able to render virtual objects with scene consistent illumination, including shadows). This chapter addresses the latter of these challenges.
Figure 1 shows how important it is to handle the dynamically changing illumination in outdoor scenarios when aiming at visually credible augmented reality (Jensen, Andersen, & Madsen, 2006). As will become apparent in the review of related work in section 2 the bulk of the work in illumination consistency for AR is based on measuring the illumination using some sort of known illumination calibration object, a probe. Obviously, for operational AR in outdoor scenes, with rapidly changing illumination, the requirement that the illumination be *captured/measured* is prohibitive.

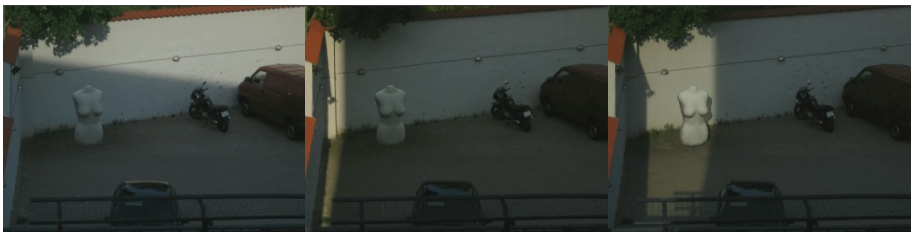

Fig. 1. Three frames from a 3 hour long sequence showing virtual sculpture rendered into scene with consistent illumination.

Source: Augmented Reality, Book edited by: Soha Maad,
ISBN 978-953-7619-69-5, pp. 230, January 2010, INTECH, Croatia, downloaded from SCIYO.COM

This chapter describes a technique we have developed which allows for illumination estimation *directly* from the images with no requirement of illumination calibration objects in the scene prior to image acquisition. This enables the approach to function adaptively in the sense that it can automatically adapt to the changing illumination conditions.
The work presented here is based on images of outdoor scenes in daylight conditions captured with a commercially available stereo camera (a Bumblebee XB3 camera from PointGrey Research, Inc.) providing dense 3D and color information of the scene. The AR renderings we show are rendered with non-realtime rendering software, but in section 8 we discuss how the technique can be implemented to operate in real-time. At present we are in fact making such an implementation.

2. Overview of related work

The classical approach to achieving illumination consistency for AR is to employ a light probe image, also called an environment map. A light probe is an omni-directional image of the scene, acquired at the location where a virtual object is to be inserted. The light probe image must be acquired in High Dynamic Range (HDR) to capture illumination sources without pixel saturation. This approach was pioneered in (Debevec, 1998) and has later been employed as the basic illumination technique in a multitude of real-time AR research (Barsi et al., 2005); (Franke & Jung, 2008); (Havran et al., 2005); (Kanbara & Yokoya, 2004); (Madsen & Laursen, 2007).
While the approach of using a captured HDR environment map of the scene allows for really realistic illumination interaction between real and virtual scene elements, it is of course totally impractical for an adaptive real-time system to require such environment maps. They must be captured where an augmentation is to be made, and the basic idea cannot be used for scenes with dynamically changing illumination.
Another family of research into AR illumination is centered around manually creating detailed 3D model of the real scene, including the sizes and positions of the light sources, combined with one or more images of the scene. Using inverse rendering techniques virtual objects can then be rendered into views of the scene with consistent illumination. Examples of work in this category include (Boivin & Gagalowitcz, 2001); (Drettakis et al., 1997); (Yu et al., 1999); (Loscos et al., 2000). The amount of manual 3D modelling required by these techniques, combined with the need for multiple views of the scene (for some of the methods), make them unrealistic for general purpose real-time illumination adaptive AR systems.
In addition to the work mentioned above there is a large body of research into using image information directly to estimate the real scene illumination conditions. Some work utilizes the shadows cast by a *known* object to compute the illumination (Sato, Sato, & Ikeuchi, CVPR 1999, 1999); (Sato, Sato, & Ikeuchi, 2001); (Kim & Hong, 2005). The disadvantage with these techniques is that it is necesary to place a known object, e.g., a cube, in the scene, and the techniques are based on an assumption that the known objects casts shadows on some known geometry in the scene, typically a planar surface.
Other work is based on estimating the illumination using shading information from images of glossy spheres in the scene (Hara et al., 2005); (Zhou & Kambhamettu, 2008). Some work combine the information from both shadows and from shading, (Wang & Samaras, 2003). Again, these techniques are impractical in the fact that they require an object in the scene for which perfect geometric information is available.

The final category of related work we mention is quite related to the research presented in this paper in the sense that it also addresses outdoor daylight scenarios. Under the assumption that two different views of a static outdoor daylight scene are available, and assuming that there are shadows from vertical structures in the scene, the camera can be calibrated to the scene and the direction vector to the sun can be estimated, (Cao & Foroosh, 2007); (Cao & Shah, 2005). In that work manual identification of essential object and shadow points is required. Shadows from virtual objects onto real scene surfaces are generated using information extracted from semi-manually detected real shadows, under the assumption that the shadows fall on uniformly colored surfaces.

As seen from the above overview of related work there is a substantial body of research into the problem of achieving illumination consistency in AR. Nevertheless, all of the presented work has some inherent drawbacks in terms of being suitable for an AR system which is to automatically adapt to changing illumination conditions. As described these drawback are things such as a requirement for known objects in the scene, the need to acquire an omni-directional HDR image the scene at the location of a virtual object, or the need for manual 3D scene modelling and/or identification of essential elements in the images, and in most cases an assumption that the scene is static.

In this chapter we present our proposed approach to dealing with these problems. The title of the chapter refers to the fact that our technique does not require placing known probes or calibration objects in the scene. The technique presented here is based on 3D scene information acquired automatically by a commercially available stereo camera, combined with illumination information extracted automatically and directly from shadow information in the input image stream, thus completely obliviating the need for manual input. The approach is suitable for real-time operation, and is based on a sound theoretical model of the problem domain.

3. Overview of approach

The input is a video stream of color images of a dynamic outdoor daylight scene. In addition to the color images we have partial 3D information of the scene in the form of a stream of depth maps generated by a stereo camera (Bumblebee XB3 from PointGrey Research, Inc.). A color image and its corresponding 3D scene data is shown in Figure 2.

Fig. 2. Left: color image from video sequence. Right: corresponding 3D scene model.

Our approach is essentially a three-step process. In the first step the combined color and depth data stream is monitored continuously and shadows from dynamic objects are

detected. Such dynamic objects can be people walking by, cars driving by, or shadows from tree leaves moving in the wind. In the second step the color information from detected shadows, combined with the depth information from the scene, are combined in an estimation of the real scene illumination. In essence the color and intensity of the sky are estimated, as are the color and intensity of the sun. In more rigorous radiometric terms we estimate the radiances (in three color channels) of the sky and the sun. In the third step, the estimated illumination information, plus the color image and the depth data, are used to render virtual objects into the video sequence.

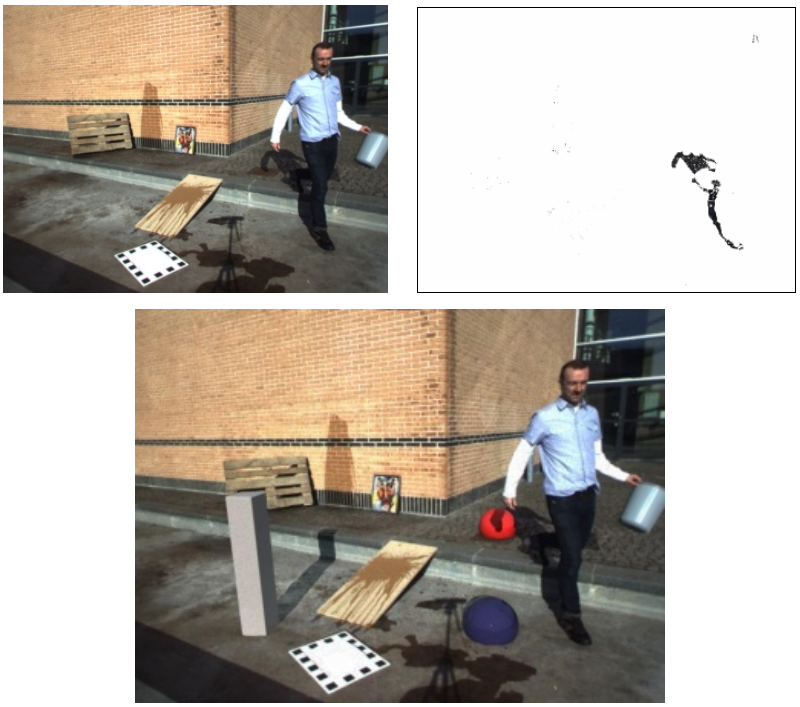

Fig. 3. Top left: input image containing walking person. Top right: detected dynamic shadow. Bottom: image augmented with virtual objects consistently shaded and casting shadows.

The position of (direction vector to) the sun is computed from additional input information. Given knowledge of the date, the time, the location on Earth (from GPS), the direction of North (from a compass) and the inclination of the camera (from an inertial sensor) the direction vector to the sun can be computed using standard expressions, (Madsen & Nielsen, 2008). Using this extra information is not considered a limitation of the approach as more and more consumer cameras already have such sensors built-in.

4. Theoretical framework

The work presented in this chapter builds upon research presented in (Madsen et al., 2009) and (Madsen & Nielsen, 2008). The novelty in this chapter lies in demonstrating how these

two previously published techniques can work together to achieve adaptive illumination for AR. The interested reader is referred to these papers for a little more depth on the theoretical framework, but a crucial element in both the shadow detection and the illumination estimation steps outlined in section 3 is a fundamental assumption that the real scene is dominated by diffuse surfaces (brick, asphalt, concrete, etc.). The reflected radiance from a diffuse surface is proportional to the incident irradiance. When taking an image of such a surface the pixel value given by the camera is proportional to the outgoing radiance from the surface at that location. The constant of proportionality depends on things such as lens geometry, shutter time, lens aperture, white balancing, camera ISO setting, etc. In practice the constant of proportionality is unknown and can only be determined through complex radiometric calibration of the camera. Our work is based on relative image measurement where this constant of proportionality cancels out and is immaterial, i.e. can remain unknown.
The pixel value, P, corresponding to a point on a diffuse surface can be formulated as:

$$P = c \cdot \rho \cdot E_i \cdot \frac{1}{\pi}$$

Here c is the aforementioned constant of proportionality, ρ is the diffuse albedo of the surface point, and E_i is the incident irradiance at the surface point. All of these variables are color channel dependent, but we just present the expression here as a general cholor channel independent expression.
Given the assumption of an outdoor, daylight scenario, the incident irradiance is the sum of the flux density received from the sky and from the sun. In other words:

$$E_i = E_{sky} + E_{sun}$$

The two-part irradiance, consisting of the irradiance from the sky and from the sun, respectively, is central to our approach. This can be described as follows. Assuming for now that we actually can detect dynamic shadows, i.e. detect when a pixel goes from being illuminated by both the sky and the sun to being illumination only by the sky, then we could divide the pixel values corresponding to the two pixel conditions (shadow, not shadow):

$$\frac{P_{shadow}}{P_{sun}} = \frac{c \cdot \rho \cdot E_{sky}}{c \cdot \rho \cdot (E_{sky} + E_{sun})} = \frac{E_{sky}}{(E_{sky} + E_{sun})}$$

That is, the ratio of pixel value in shadow over the pixel value in sun is indepedent on camera constants and surface albedo, and only depends on the shadow to sun irradiance ratios. Thus the detected dynamic shadows contain essential illumination information that can be used to compute the radiances of the sky and the sun, respectively.
For a given surface point a number of factors determine exactly what amount of irradiance is received from the sky and the sun, respectively. These factors relate to the scene geometry, i.e. the angle between the surface point normal and the direction vector to the sun, and the amount of sky dome visible from the surface point (parts of the sky dome are occluded by scene geometry such as walls). For the sun contribution these elements can be formulated as follows:

$$E_{sun} = (\vec{n} \cdot \vec{s}) \cdot E_{sun}^{\uparrow}$$

The normal vector of the surface is $\vec{n}$, the direction vector to the sun is $\vec{s}$, and $E_{sun}^{\uparrow}$ is the irradiance received from the sun by surface perpendicular to the direction vector to the sun (surface normal pointing straight into the sun). The direction vector to the sun is computed based on GPS and date/time information, as described above, and the normal information per surface point comes from the depth data.
For the sky contribution the geometric factors can be formulated as:

$$E_{sky} = V_{sky} \cdot E_{sky}^{\uparrow}$$

where $E_{sky}^{\uparrow}$ is the irradiance from the sky by a surface point which „can see" the whole sky dome, and V_{sky} is the fraction (range 0 to 1) of the sky dome a given surface point can see (ambient occlusion term). The ambient occlusion term can be computed per surface point from the depth data of the scene.
Our task thus is to estimate the two terms, $E_{sky}^{\uparrow}$ and $E_{sun}^{\uparrow}$, using the detected dynamic shadows, combined with per surface point geometrical information from the depth data. Given that we can estimate these two irradiance contributions we have all the information required to render virtual objects into the scene with proper shadows and shading.

5. Detection of shadows

The first step in our approach is to detect pixels that represent shadows cast by objects that are moving in the scene. Dynamic shadow detection based on a combination of color and depth information is presented in detail in (Madsen, Moeslund, Pal, & Balasubramanian, 2009). In this section we give an overview of the technique. First we give a more precise description of the equipment and setup used for the experimental results.

5.1 Equipment and setup

We are aiming a constructing a real-time AR setup with a computer monitor mounted on a pole so that a user can rotate the monitor around a vertical and a horizontal axis, essentially being able to aim the monitor in all direction, see Figure 4.

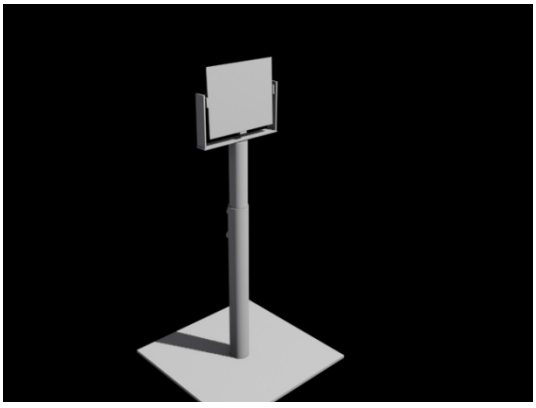

Fig. 4. 3D model of a computer monitor mounted on pole to allow users to aim monitor in various directions.

Behind the monitor a camera will be placed, and the monitor will show the video image from that camera, including augmented virtual objects. For the experiments presented in this chapter we use a static camera.
As described the camera is a commercially available stereo camera, the Bumblebee XB3 from PointGrey Research. This camera provides rectified color images, and the accompanying API enables generation of dense disparity maps, which can be converted into metric depth maps. On an Intel Core Duo 2 2.3 GHz maching with 2 Gbyte memory, 640x480 disparity maps can be generated at 10 frames per second. We operate the disparity generation process in a mode which provides sub-pixel disparities encoded in 16 bit resolution to have as high as possible depth precision. The color images are also in 640x480 pixel resolution. For all images shown in this chapter we have used to rightmost of the two color images supplied by the stereo camera.
We have used Google Maps to look up the earth latitude and longitude of the positions where we have placed the camera for the reported test sequences. The camera is calibrated to a scene based coordinate system using a standard camera calibration process involving a checkerboard pattern. This pattern is visible in the images shown in Figure 2. We orient the pattern such that one of the axes is pointing North (done with a compas).
In the future when we place the stereo camera on the pan-tilt rig shown in Figure 4, we will use angle encoders to determine the pose (viewing direction) of the camera in real-time, and then the orientation only has to be initialized to level and North in an initialization phase, i.e. leveling the camera and pointing it due North and pressing a key on the computer. From then on the system will know the camera pose with very high precision, similar to what we did in (Madsen & Laursen, 2007) for a system which was not adpative to scene illumination, and which used a pre-recorded environment map for real-time shading.

5.2 Detection of shadow candidate regions

We employ a conceptually simple approach to detecting pixels that are candidates for representing a point that has come into shadow from being in direct sunlight. Dynamic shadows are traditionally detected in video sequences by employing a trained background model, see (Madsen, Moeslund, Pal, & Balasubramanian, 2009) for a review of the research literature. For our application we consider this approach to be unrealistic since we are aiming at a system which must operate for hours, days, or even more. With the drastic illumination variation that can occur over such a time span, including seasonal changes in the color and amount of vegetation we do not consider a background model as viable.
Instead, we have based our dynamic shadow detection on image differencing. To be able to respond correctly to rapid illumination changes on windy, partly overcast days, we continuously subtract a 0.5 second old frame from the current frame. This is done both on the color image stream, and on the disparity image stream. By thresholding these difference images we get two "change has occurred" triggers, corresponding to a color change and a depth change for any given pixel, Figure 5. In this illustration we have used not a 0.5 second delay for the differencing but a 10 second delay to show changes more clearly. With a 10 second delay the system will not be as responsive to rapid illumination changes, though.
Regions where there is no depth change, but where there is a color change, represent shadow candidate pixels. In the example in Figure 5 the walking person has poured water on to surfaces in the scene, effectively changing the surface albedo at those locations. In this first step these regions qualify as shadow regions, but are discarded in a subsequent step.

Fig. 5. Left: thresholded disparity difference (change in depth). Center: thresholded color difference (change in color). Right: changes classified into moving object (grey) and shadow candidates.

5.3 Verification of shadow regions

The shadow candidate pixels detected by the color/depth differencing step are analyzed for how well each pixel conforms to the expected behavior of a pixel coming into shadow in an outdoor daylight scene. Given the combined sky and sun illumination environment, and given the fact that the sky is much more bluish than the sun, shadows undergo a color transformation which can be termed a blue shift. The chromaticity (red and blue channels normalized with the green channel) of a pixel will shift towards blue, when a pixel transitions from being in direct sunlight to being in shadow; this is proven in (Madsen, Moeslund, Pal, & Balasubramanian, 2009).

This can be exploited to filter out those shadow candidates which are not really shadows, such as the water splashes in Fig. 5. This is considered a powerful feature of our approach since we do not have to worry that a sudden rain shower will cause false detections of shadow regions.

We have designed a simple filter which only accepts those shadow candidate pixels which vote for the same blue shift direction in chromaticity space. The accepted shadow pixels are shown in Fig. 6, which clearly demonstrate that water splashes are not detected as shadow pixels.

Fig. 6. Input image and verified dynamic shadow pixels.

6. Illumination estimation

Having thus demonstrated how dynamic shadows can be detected in the scenes, time has come to look at how these shadows can be used to estimate the illumination conditions, or

more precisely: estimate the radiances (in three color channels) of the sky and the sun, respectively.

6.1 Assumptions

The illumination estimation is described in greater detail in (Madsen & Nielsen, 2008). In this chapter we focus a little more on the overall approach, rather than going into detail with the derivation of the expressions. Yet, the process rests on a small number of assumptions that are worth mentioning.

First of all it is assumed that a majority of the surfaces in the scene behave as Lambertian reflectors, i.e. are diffuse and reflect light equally in all directions. In our experience ordinary outdoor man-made environments exhibit such behavior quite well, for example materials such as concrete, brick and asphalt. At normal imaging distances grass areas are also approximately diffuse (when getting too close the individual grass straws can be quite glossy).

Furthermore, as described, it is assumed that the only illuminants are the sky and the sun. Artificial light sources, such as outdoor lamps at night or near-night conditions are not modeled and we cannot at present, with the framework presented here, take their contribution into account. Strong solar reflections from windows in buildings can also present an illumination contribution which is not part of the model. If such reflections are significantly present in the scene, the contribution will be included in the estimate of the irradiance from the sky dome.

We also assume that we have all the information available to use an analytical approach to determining the direction vector to the sun (as described in section 3).

Finally, it is assumed that the camera is white balanced such that if a white/grey surface were placed at a known orientation, for example straight up, in the scene it would actually image as a white-balanced region. Notice that we do not assume such a surface to be *present* in the scene. The estimated illumination (sky and sun radiances) using our proposed approach will be tuned such that a white virtual surface with normal pointing up will render to a white-balanced surface. This way we can operate the camera with automatic white-balancing and thus maintain proper color balances across day long sequences, and across various weather conditions (overcast to bright sunlight).

6.2 From detected shadows to estimated illumination model

Section 4 presented the theoretical framework for formulating the relationships between measured pixels values, sun to shadow ratios and the various geometric properties of the scene. Key to the illumination estimation process is per pixel information about normal direction and ambient occlusion term. These are easy to compute from the depth data provided by the stereo camera, see Figure 7. At present we compute per pixel ambient occlusion by using a ray tracer (RADIANCE) to shade the geometry with a sky dome having a radiance of $1/\pi$ [W/(m² sr)]. With such a radiance any point receiving light from the entire sky dome will receive an irradiance of 1 [W/m²], and a point receiving light from half the dome will receive an irradiance of 0.5 [W/m²], etc.

With this geometrical information, combined with knowledge of the direction vector to the sun, each detected and verified shadow pixel can essentially vote for the two unknown scene illuminants, namely the unknown sky and sun irradiances, $E_{sky}^{\uparrow}$ and $E_{sun}^{\uparrow}$, as described in section 4.

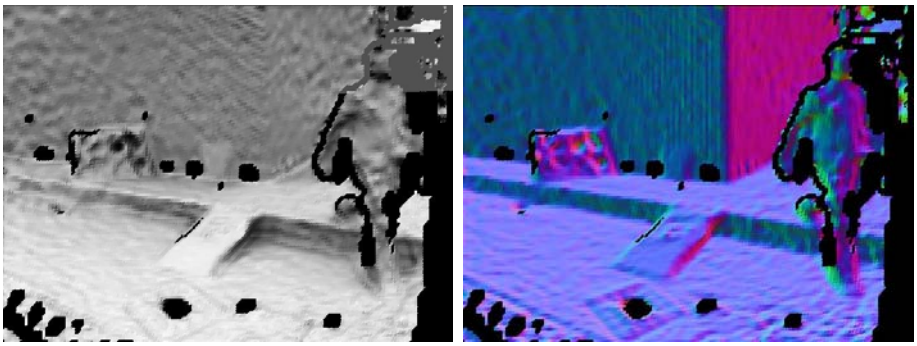

Fig. 7. Left: ambient occlusion coded as grey level in the range from 0 to 1, where darker areas see a smaller fraction of the sky dome. Right: per pixel surface normals shown as a normal map with xyz normal components encoded into offset, normalized rgb values.

In theory a single shadow pixel is enough to provide information about the sky and sun irradiances, but naturally we achieve a much more robust estimate by incorporating all shadow pixel into a voting scheme to find the sky and sun irradiances which have the strongest support in the available data. This way we can also filter out outliers due to any pixels falsely labelled as shadow, and thus we are less sensitive to any violation of the assumption that surfaces are diffuse.

For the frame shown in Figure 7 the approach estimates the radiance of the sky to [0.72 0.79 0.87] (rgb values, respectively) and the sun radiance to [69000 64000 59000]. We convert the estimated irradiances into radiances, since the radiances are more suitable for actual rendering. Radiance is the proper measure of the „brightness" of a light source. We notice a couple of things from these estimated numbers. First of all the color balance of the sky is towards blue as compared to the sun's rgb values. Secondly, there is roughly five orders of magnitude difference between the sky radiance and the sun radiance, which is entirely consistent with what is expected. As a sule of thumb the combined irradiance from the sky is considered to be equal to the irradiance from the sun, but the sun is 5 orders of magnitude smaller than the sky, and therefore 5 orders of magnitude brighter (5 orders of magnitude higher radiance).

7. Rendering of augmentations

After the radiances (in all three color channels) of both the sky and the sun have been estimated it is possible to setup the rendering of augmented objects. Available for this process are: the color image of the scene, the 3D model of the scene (the depth map), and the illumination environment.

Specifically, the illumination environment is set up as a hemi-spherical light source (hemi-spherical around the up-vector of the scene), combined with a disc source to act as the sun. The sun disk source has a 0.53 degree diameter, and the direction vector to the disk source is the computed sun position corresponding to the time and place of the image acquisition. The hemi-spherical and the disk light sources are assigned the estimated radiances, respectively.

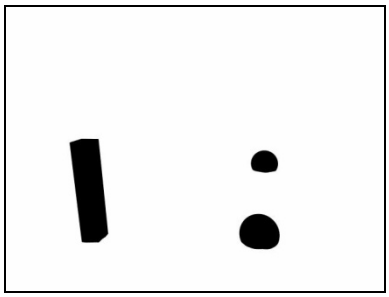

Fig. 8. Binary mask indicating image positions of augmented objects.

For the renderings demonstrated in the chapter we have employed the RADIANCE rendering package. The depth map is imported as a triangle mesh, and augmented objects are modelled as triangle meshes or volume primitives into the same coordinate system as the mesh.

We employ a basic technique similar to the differential rendering approach described in (Debevec, 1998). Roughly speaking this technique focuses on rendering the modelled part of the scene *with* augmentation, and rendering it *without*, finding the difference between these renderings and adding the difference to the original image (hence the name differential rendering).

First, though, the albedos of the real surfaces in the scene are estimated. This is done iteratively as follows: the input image is copied to an new image called the albedo map, and the scene mesh from the depth map is textured with this albedo map. Then the scene is rendered with the estimated illumination environment and the rendered image is divided into the input image, yielding a scale map. This scale map is multiplied with the first version of the albedo map to produce a new albedo map; using this map as mesh texture, the scene is rendered with the estimated illumination again, and the result is divided into the input image. This process is continued until there are no significant changes in the albedo map. Typically, three to four iterations suffice. This process will eliminate all global illumination effects from the albedo map. All subsequent rendering involving the real scene geometry uses this estimated albedo map as texture for the real scene mesh.

Then a binary mask image is rendered to indicate where in the image the augmented object will be projected, see Figure 8. This mask is used to mask out areas in the image resulting from rendering the scene mesh, textured with estimated albedo map, and illuminated with the estimated light. Then the scene mesh, textured with the albedo map, *plus* the augmented objects, are rendered with the same illumination. These two renderings are subtracted, see Figure 9. The shadows cast by the augmented objects will represent negative pixel values in this differential image, which is added to the masked input image to produce the final render, see Figure 3.

The described approach has been applied to multiple frames from a range of different input sequences acquired at different times of day, and under varying weather conditions, and some examples are shown in Figure 10. Since we have an estimated radiance for an entire sky dome (hemi-sphere) we can allow the scene to contain quite glossy objects, which can then reflect the illumination from the sky dome. The color balance of the estimated sky radiances is very consistent with what is expected from looking at the images (which to some extent indicate the sky color even if sky pixels are a close to being saturated). Naturally, with the technique presented here, the entire sky dome will have uniform color,

and any illumination effect in the real scene from clouds will sort of be spread across the entire estimated sky. Therefore, if the virtual objects augmented into the scene contains too glossy surfaces, where one would expect to see reflections of clouds (if any in the real scene), the actual sky reflections will be uniformly colored and cloud-less.

Fig. 9. Modelled part of scene rendered with no augmentation and masked with binary mask indicating augmented objects (left), is subtracted from scene rendered with augmented objects (center), yielding a difference image (right).

The depth and color balance of the shadows cast by virtual objects onto real objects will more or less by definition be correct. This is because the whole basis for the estimation of the illumination environment is the chromatic properties of the detected *real* shadows in the scene.

The 3D scene information from the stereo camera naturally also contains information about the dynamic objects in the scene. In the example images presented in this chapter the primary author is present to provide movements generating shadows for detection and illumination estimation. The 3D mesh information concerning him enables various shadow interaction between real and virtual geometry. In Figure 9 it can be seen that the hand of the walking person casts a shadow on the red augmented sphere; in Figure 10 (top) the entire red sphere in the front of the scene is in the shadow cast by him; in the same figure (bottom) the person's lower legs are partially occluded by the virtual pillars inserted into the scene.

8. Discussions and future work

The entire shadow detection and illumination estimation can run at interactive frame rates (approx. 8 frames per second) in our present C++ implementation. The rendering process described in section 7 is at present entirely non-real-time since it is based on physically-based ray-tracing. Nevertheless, we have previously demonstrated that albedo estimation as well as rendering augmentations into real scene with complex shadow interaction can be done in real-time, (Madsen & Laursen, 2007). The hemi-spherical sky dome illumination can be replaced with a set of directional lights scattered uniformly across a hemi-sphere, and the sun can be rendered as a single directional light source.

A key element behind the technique presented here is an assumption that the real scene is dominated by surfaces which approximately act as diffuse reflectors (follow Lambert's cosine law). As described we have experienced this to be a fair assumption in typical man-made urban environments. We are presently investigating techniques based on robust statistics to identify areas in the scene, which do not over long sequences concur with the rest of the scene in terms of estimated illumination. Since we get illumination estimation per

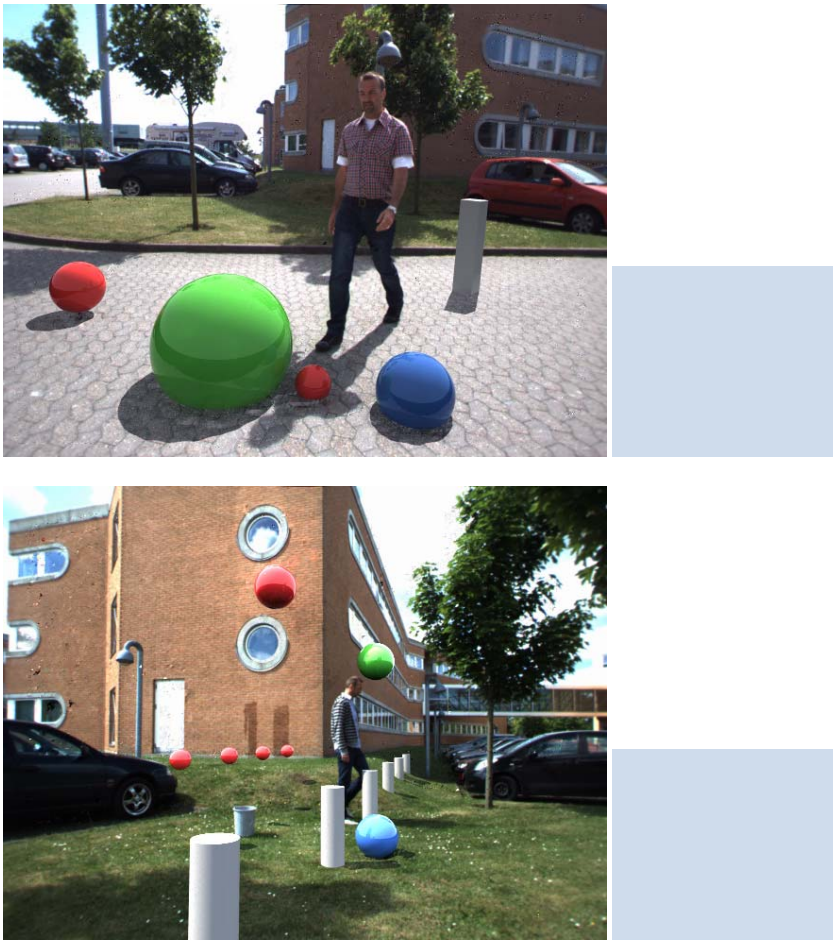

Fig. 10. Two different augmented scenes. The small blue fields on the right are the estimated sky colors (the color balance estimated for the sky using the presented technique).

pixel and use voting schemes it is actually possible to label the pixels which do not exhibit a temporal color development which is consistent with the illumination conditions voted for by the majority of the pixels. This will enable us to label and rule-out highly glossy surfaces such as windows, pools of water, and metallic surfaces such as cars.

At present our technique cannot handle the problem of double shadows (cases where a shadow cast by a virtual object falls on top of an already existing shadow in the scene, creating a double shadow). If the 3D data from the stereo camera correctly models the geometry casting the real shadow, the rendering process will ensure that no double shadows are created. For example, the little red sphere in the front of the bottom scene in Figure 10 does not cast a double shadow inside the shadow already cast by the walking person. But, if the shadow caster geometry is not available, e.g. is outside the camera's field-of-view, double shadows can occur. We are developing techniques to deal with this problem

based on creating "shadow protection masks" from the detected shadows in the scene. Since we know where the real shadows are we can use these masks to avoid casting extra shadow in them.

9. Conclusion

For photo-realistic Augmented Reality it is important to ensure that the augmented objects are rendered with illumination which is consistent with the real scene illumination. Especially the shadows are very important. Related work on illumination consistency for AR all involves a special purpose, known object to be placed in the scene, or requires manual interaction. In this chapter we have presented a novel, completely alternative, technique which does not have these constraints. Using only real-time stereo images as input the proposed technique can estimate the illumination information for an outdoor scene directly from image measurements of natural scenes.

An important additional benefit of the presented approach is that it does not require High Dynamic Range imagery, and the technique allows the system to use automatic gain control (AGC) on the camera, so that the system can operate across very long sequences with proper exposure of the images. In fact the estimated illumination parameters will follow the AGC on the camera so that, if the camera reduces light sensitivity with for example 10%, the estimated illumination is also 10% lower. This means that the rendered virtual objects will appear directly to be subject to the same AGC behaviour and thus match the real scene images perfectly. The same argumentation is valid for automatic camera white-balancing (if present).

With this work we believe we are an important step closer to being able to make Augmented Reality systems that function in real-time, and automatically adapt seamlessly to changes in the illumination conditions in the scene.

10. Acknowledgements

This research is funded by the CoSPE (26-04-0171) project under the Danish Research Agency. This support is gratefully acknowledged.

11. References

Barsi, A., Szimary-Kalos, L., & Szecsi, L. (2005). Image-based Illumination on the GPU. *Machine Graphics and Vision* , 159 - 169.

Boivin, S., & Gagalowitcz, A. (2001). Image-Based Rendering of Diffuse, Specular and Glossy Surfaces from a Single Image. *Proceedings: SIGGRAPH.* Los Angeles, California.

Cao, X., & Foroosh, H. (2007). Camera Calibration and Light Source Orientation from Solar Shadows. *Computer Vision and Image Understanding* , 60 - 72.

Cao, X., & Shah, M. (2005). Camera Calibration and Light Source Estimation from Images with Shadows. *Proceedings: IEEE Conference on Computer Vision and Pattern Recognition*, (s. 928 - 923).

Debevec, P. (1998). Rendering Synthetic Objects into Real Scenes: Bridging Traditional and Image-based Graphics with Global Illumination and High Dynamic Range Photography. *SIGGRAPH*. Orlando, Florida.

Drettakis, G., Robert, L., & Bougnoux, S. (1997). Interactive Common Illumination for Computer Augmented Reality. *Proceedings: Eurographics*. St. Etienne, France.

Franke, T., & Jung, Y. (2008). Real-Time Mixed Reality with GPU Techniques. *Proceedings: International Conference on Graphics Theory and Applications*. Funchal, Maderia, Portugal.

Hara, K., Nishino, K., & Ikeuchi, K. (2005). Light Source Position and Reflectance Estimation from a Single View withou the Distant Illumination Assumption. *IEEE Transactions on Pattern Analysis and Machine Intelligence* , 493 - 5050.

Havran, V., Smyk, M., Krawczyk, G., Myszkowski, K., & Seidel, H.-P. (2005). Importance Sampling for Video Environment Maps. *Proceedings: Eurographics Symposium on Rendering*, (s. 31 - 42). Konstanz, Germany.

Jensen, T., Andersen, M., & Madsen, C. B. (2006). Real-time Image-Based Lighting for Outdoor Augmented Reality Under Dynamically Changing Illumination Conditions. *Proceedings: International Conference on Graphics Theory and Applications* (pp. 252 - 261). Setubal, Portugal: INSTIC.

Kanbara, M., & Yokoya, N. (2004). Real-Time Estimation of Light Source Environment for Photorealistic Augmented Reality. *Proceedings: International Conference on Pattern Recognition*, (s. 911 - 914). Cambridge, United Kingdom.

Kim, T., & Hong, K.-I. (2005). A Practical Single Image Based Approach for Estimating Illumination Distribution from Shadows. *Proceedings: IEEE International Conference on Computer Vision*, (s. 266 - 271).

Loscos, C., Drettakis, G., & Robert, L. (2000). Interactive Virtual Relighting of Real Scenes. *IEEE Transactions on Visualization and Computer Graphics , 6* (4), 289 - 305.

Madsen, C. B., & Laursen, R. E. (2007). A Scalable GPU-Based Approch to Shading and Shadowing for Photorealistic Real-Time Augmented Reality. *Proceedings: International Conference on Graphics Theory and Applications*, (pp. 252 - 261). Barcelona, Spain.

Madsen, C. B., & Nielsen, M. (2008). Towards Probe-less Augmented Reality - a Position Paper. *Proceedings: International Conference on Graphics Theory and Applications*, (s. 255 - 261). Funchal, Madeira, Portugal.

Madsen, C. B., Moeslund, T. B., Pal, A., & Balasubramanian, S. (2009). Shadow Detection in Dynamic Scenes Using Dense Stereo and an Outdoor Illumination Model. In R. Koch, & A. Kold (Ed.), *Dynamic 3D Imaging. LNCS 5742*, pp. 110 - 125. Berlin Heidelberg: Springer-Verlag.

Sato, I., Sato, Y., & Ikeuchi, K. (1999). Illumination Estimation from Shadows. *Proceedings: IEEE Conference on Computer Vision and Pattern Recognition*, (s. 306 - 312).

Sato, I., Sato, Y., & Ikeuchi, K. (2001). Stability Issues in Recovering Illumination Distribution from Brightness in Shadows. *Proceedings: IEEE Conference on Computer Vision and Pattern Recognition*, (s. 400 - 407).

Wang, Y., & Samaras, D. (2003). Estimation of Multiple Directional Light Sources for Synthesis of Augmented Reality Images. *Graphical Models , 65*, 185 - 205.

Yu, Y., Debevec, P., Malik, J., & Hawkins, T. (1999). Inverse Global Illumination: Recovering Reflectance Models of Real Scenes From Photographs. *Proceedings: SIGRRAPH*, (s. 215 - 224). Los Angeles, California.

Zhou, W., & Kambhamettu, C. (2008). A Unified Framework for Scene Illumination Estimation. *Image and Vision Computing* , 415 - 429.